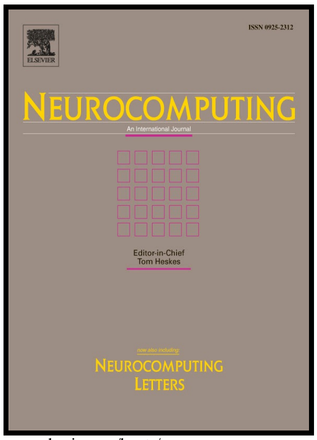
Author's Accepted Manuscript



www.elsevier.com/locate/neucom

Improved SFFS method for channel selection in motor imagery based BCI

Zhaoyang Qiu, Jing Jin, Hak-Keung Lam, Yu Zhang, Xingyu Wang, Andrzej Cichocki

PII: S0925-2312(16)30374-5
DOI: http://dx.doi.org/10.1016/j.neucom.2016.05.035
Reference: NEUCOM17052

To appear in: *Neurocomputing*

Received date: 1 December 2015
Revised date: 18 March 2016
Accepted date: 12 May 2016

Improved SFFS method for channel selection in motor imagery based BCI

Zhaoyang Qiu[1], Jing Jin[1], Hak-Keung Lam[2], Yu Zhang[1], Xingyu Wang[1], Andrzej Cichocki[3]

Abstract

Background: Multichannels used in brain–computer interface (BCI) systems contain redundant information and cause inconvenience for practical application. Channel selection can enhance the performance of BCI by removing task-irrelevant and redundant channels. Sequential floating forward selection (SFFS) is an intelligent search algorithm and is considered one of the best feature selection methods in the literature. However, SFFS is time consuming when the number of features is large.

Method: In this study, the SFFS method was improved to select channels for the common spatial pattern (CSP) in motor imagery (MI)-based BCI. Based on the distribution of channels in the cerebral cortex, the adjacent channels would be treated as one feature for selection. Thus, in the search process, the improved SFFS could select or remove several channels in each iteration and reduce the total computation time.

Results: The improved SFFS yielded significantly better performance than using all channels ($p < 0.01$) and support vector machine recursive feature elimination method ($p < 0.05$). The computation time of the proposed method was significantly reduced ($p < 0.005$) compared with the original SFFS method.

Conclusions: This study improved the SFFS method to select channels for CSP in MI-based BCI. The improved SFFS method could significantly reduce computation time compared with the original SFFS without compromising the classification accuracy. This study provided a way to optimize electroencephalogram channels, which combined the distribution of channels and the intelligent selection method (SFFS). Improvements were mainly in the perspective of reducing computation time, which leads to convenience in the practical application of BCI systems.

Keywords: brain–computer interface (BCI), motor imagery, channels selection, SFFS

1. Introduction

Brain–computer interface (BCI) systems can translate brain activities into commands used to control external devices [1, 2]. BCI systems provide a new communication method for people with severe neuromuscular disabilities, such as amyotrophic lateral sclerosis, spinal cord injury, and traumatic brain injury [3, 4]. Various neural activities can be used as features in electroencephalogram (EEG)-based BCI systems. P300 evoked potentials [5–10], slow cortical potentials, steady state visually evoked potentials [11, 12], and event-related desynchronization (ERD) [13] / event-related synchronization

Jing Jin (✉) jinjingat@gmail.com

Xingyu Wang (✉) xywang@ecust.edu.cn

1. Key Laboratory of Advanced Control and Optimization for Chemical Processes, Ministry of Education, East China University of Science and Technology, Shanghai, P.R. China

2. Department of Informatics, King's College London, London WC2R 2LS, UK.

3. Laboratory for Advanced Brain Signal Processing, Brain Science Institute, RIKEN, Wako-shi, 351-0198, Japan and Systems Research Institute of Polish Academy of Science, Warsaw, Poland

(ERS) [14] are extensively used in BCI systems.

Motor imagery (MI)-based BCI systems work without external stimulus and are operated more easily compared with stimuli-based BCI [15, 16]. Most BCI systems require multichannel EEG data to achieve good performance [17]. However, many channels would contain redundant information and noise for data processing and cause inconvenience for practical application [18–20]. In several cases, no clear agreement exists on the number and location of necessary channels for MI [21]. Thus, channel selection methods are necessary to improve the performance of MI-based BCI.

Various channel selection methods have been proposed [22–25]. In several studies, channels were selected manually based on neurophysiologic knowledge. MI-based BCI commonly uses C3, C4, and Cz channels, which record important characteristics of MI [26–29]. Wrapper-based methods, filter-based methods, and common spatial pattern (CSP)-based methods are used to select EEG channels more often than the manual selection method [30]. Wrapper-based methods usually select channels coupled with a specific classifier [20, 31]; its performance mainly depends on the applied classifier. Filter-based channel ranking methods are commonly found in the literature, such as the mutual information-based channel selection method [32, 33] and the Fisher score-based channel selection method [34]. In these two filter-based methods, channels are ranked individually. In general, good features do not necessarily lead to good performance [35]. As a method for feature extraction, CSP is effective in detecting MI EEG signals [36–39]. The CSP algorithm has also shown validity in channel selection [40, 41]. In [42], the regularized CSP was used to select channels.

Typically, channel selection could work toward two opposite directions: 1) to select the most effective channels one by one and 2) to eliminate noisy channels one by one [30]. The sequential floating forward selection (SFFS) was originally developed by Pudil et al. [43]. SFFS takes the most significant feature from the remaining features at each time, inserts it into the selected feature subset, and dynamically deletes the most meaningless feature from the selected feature subset. Thus, the SFFS is appropriate for channel selection. However, this method is time consuming, particularly when the number of features is large. To address this issue, the present study proposes the improved SFFS to select the optimal channels. The improved SFFS could select or delete several channels at each time. Two datasets from publicly available BCI competitions were used to evaluate the performance of the improved SFFS algorithm: one dataset with a moderate number of initial channels (59 channels) and the other dataset with denser electrodes (118 channels). The classification performance of the improved SFFS was also compared with the support vector machine recursive feature elimination (SVM-RFE) method [44].

The paper is structured as follows: Section 2 describes the applied datasets and the proposed method. Section 3 shows the results of computation time and the classification accuracies. Section 4 presents the discussion. Finally, Section 5 concludes this study.

2. Methods

2.1. Description of the data

Data 1 (BCI Competition IV datasets 1): This dataset was recorded from healthy subjects at 59 EEG channels. In the entire session, MI was performed without feedback. For each subject, two classes of MI were selected from the following three classes: left hand, right hand, and foot. In each run, visual cues on a computer screen were in the form of arrows pointing left, right, or down. Cues were displayed for a period of 4 s, during which the subject was instructed to perform the corresponding MI task. These periods were interleaved with 2 s of blank screen and 2 s of a fixation cross shown at the center of the screen. The fixation cross was superimposed on the cues, i.e., it was shown for 6 s. The calibration data of this dataset include two runs with each run consisting of 100 single trials. The sampling rate is 100 Hz. In the present study, the first run was selected as the training data and the second run was selected as the test data. The experimental process of a trial is illustrated in Fig. 1. More details about the dataset can be found in the following website: http://www.bbci.de/competition/IV/.

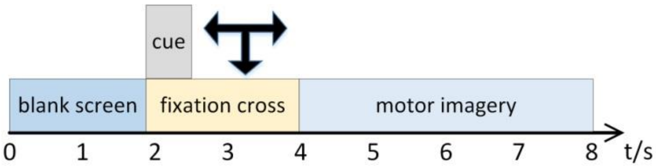


Fig. 1. Timing of a trial of data 1. Time slice between seconds 4 and 7 was used for feature extraction.

Data 2 (BCI Competition III datasets IVa): This dataset was recorded from 5 subjects at 118 EEG channels. Visual cues were displayed for 3.5 s, during which the subject had to perform the MI tasks: left hand, right hand, and foot. Only cues for the classes "right hand" and "foot" were provided for the competition. A total of 280 trials of EEG measurements for each subject are available (downsampled to 100 Hz). In the present study, the first 140 trials were selected as the training data and the remainder was utilized as the test data. The experimental process is illustrated in Fig. 2. More details about the dataset can be found in the following website: http://www.bbci.de/competition/iii/.

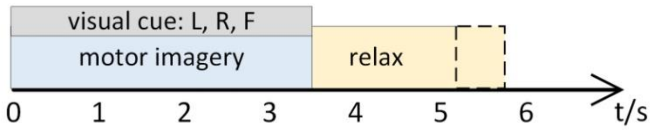


Fig. 2. Timing of a trial of data 2. Time slice between seconds 0 and 3.5 was used for feature extraction.

2.2. Common spatial pattern

The CSP algorithm is an efficient feature extraction algorithm that has been extensively used in MI-based BCI systems [45–50]. The CSP is based on the simultaneous diagonalization of two covariance matrices. This algorithm finds a spatial filter to maximize the variance for one class and minimize the variance for another class at the same time to achieve the purpose of classification. Using CPS, features were extracted from two classes of original EEG samples. A classifier can be trained with these extracted CSP features and corresponding labels.

For the analysis, the original EEG signal data are represented as a matrix $E \in R^{N \times T}$, where E is the EEG samples of a trial with dimensions $N \times T$, N is the number of channels, and T is the number of sampling points for each channel. The CSP operation process is as follows:

Calculate the spatial covariance of the EEG data:

$$C = (EE^T)/(tr(EE^T)). \tag{1}$$

C_l and C_r represent two spatial covariance matrixes (two classes of MI) , which can be calculated by averaging over the trials of each group. The composite spatial covariance matrix can be expressed as follows:

$$C_c = \overline{C_l} + \overline{C_r}. \tag{2}$$

C_c can be decomposed as follows:

$$C_c = U_c \lambda_c U_c^T, \tag{3}$$

where U_c is the matrix of eigenvectors and λ_c is the diagonal matrix of eigenvalues. In the process, the eigenvalues are arranged in descending order.

The whitening transformation is expressed as follows:

$$P = \sqrt{\lambda_c^{-1}} U_c^T. \tag{4}$$

Then, the covariance matrices $\overline{C_l}$ and $\overline{C_r}$ can be transformed into the following expressions:

$$S_l = P\overline{C_l}P^T \quad , \; S_r = P\overline{C_r}P^T \quad . \tag{5}$$

S_l and S_r share the same eigenvectors. If $S_l = B\lambda_r B^T$, then:

$$S_r = B\lambda_r B^T, \; \lambda_l + \lambda_r = I, \tag{6}$$

where I is the identity matrix. The projection matrix is achieved by the following equation:

$$F = (B^T P)^T. \tag{7}$$

This equation is the expected spatial filter. After whitening, the EEG signals can be projected on the first m and last m columns of B. Thus, the EEG data of a single trial can be transformed into the following expression:

$$Z = FE. \tag{8}$$

The present study selected m = 1. f_p can be obtained from Z_p $(p = 1 \ldots 2m)$ as the features of the original EEG data, expressed as follows:

$$f_p = \log\left(\frac{Var\left(Z_p\right)}{\sum_{i=1}^{2m} Var\left(Z_i\right)}\right). \tag{9}$$

2.3. Support vector machine

Support vector machine (SVM) is a machine learning method proposed by Vapnik et al. in the 1990s [51]. The SVM is mainly proposed for two classes of pattern recognition problems. If a given data $A \in R^d$ from two classes can be divided linearly by a hyperplane, then the hyperplane can be expressed as $WA + b = 0$. $W \in R^d$ is the weight vector and b is the intercept (scalar). Thus, the problem is transformed to find the optimal hyperplane, as follows:

$$\begin{gathered} \min \varnothing\left(W, \varepsilon\right) = \frac{1}{2} \| W \|^2 + c\sum_{i=1}^{n} \varepsilon_i, c \geq 0 \\ s.t.\; y_i (W^T A^{(i)} + b) \geq 1 - \varepsilon_i, \varepsilon_i \geq 0, i = (1, \cdots n) \end{gathered} . \tag{10}$$

The optimization problem is a convex quadratic programming problem [52], where $A^{(i)}$ is a feature vector of a training sample and y_i is the category with labels {−1, 1} in which $A^{(i)}$ belongs to. W is the hyperplane coefficient vector. The parameter ε_i is called the slack variable, and c is the regularization parameter. c is used to control the trade-off between model complexity and empirical risk [53].

For a nonlinearly separable classification problem, kernel functions can be used to map the training samples into a high-dimensional space. Thus, the resulting SVM classifier function can be rewritten as follows:

$$f\left(x\right)=\mathrm{sign}[\sum_{i=1}^{n}\alpha_i y_i K\left(a_i,a\right)+b], \tag{11}$$

where n is the number of support vectors, α_i is the Lagrange multiplier, and $K\left(a_i,a\right)$ is the kernel function. Linear kernel, polynomial kernel, and radial basis function (RBF) kernel are commonly used kernel functions. The present study used the RBF kernel, expressed as follows:

$$K\left(a_i,a\right)=\exp(-\gamma\|a_\mathrm{i}-a\|^2). \tag{12}$$

2.4. Improved sequential floating forward selection

The SFFS is a common feature selection algorithm, which is based on a bottom-up approach [43]. The SFFS is a suitable method to select EEG channels [34]. Starting from X_k, the SFFS performs the loop of channel selection continuously. In the present study, the symbol X_k denotes the channel subset which contains k channels. Y denotes the universal channel set. $J\left(X_k\right)$ denotes the performance (the cross-validation accuracy of the training data) of subset X_k.

The search process does not stop until a stopping criterion is fulfilled. This criterion can be as follows: 1) the performance of the current subset meets the requirement or 2) the desired number of channels is reached. The present study used the second criterion. When k varied from 1 to *n* (to find the optimal channels, *n* is the number of initial channels), the configuration of channels was recorded each time. Thus, *n* configurations of channels were obtained in the end: $X_1, X_2, X_3, \ldots, X_n$. The configuration with the best performance is selected as the final configuration.

If *k* features are selected from universial set Y to form the subset X_k, $J_{\max} = J(X_k)$, Fig. 3 shows the search process of the SFFS algorithm.

While the stop criterion has not been fulfilled {

Step 1 (Inclusion):

$X^+ = \mathrm{argmax}_{c\in(Y-X_k)} J(X_k + c)$ // find the significant feature with respect to the set X_k

$X_{k+1} = X_k + x^+$ // add it to X_k

$k = k + 1$

$J_{\max} = \max(J(X_{k+1}), J_{\max})$ //get the best performance of subset

Step 2 (exclusion):

$x^- = \mathrm{argmax}_{c\in X_k} J(X_k - c)$ //Find the least significant feature in the set X_k

If

$J(X_k - x^-) > J_{\max}$ // judge whether a better subset can be obtained

Then

$J_{\max} = J(X_k - x^-)$

$X_{k-1} = X_k - x^-$ //exclude the feature

$k = k - 1$

go to Step 2

Else

go to Step 1 }

Fig. 3. The process of the SFFS algorithm.

Problem: In contrast to channel ranking methods that are computationally fast, the SFFS performs the loop of channel selection continuously. The SFFS is a time-consuming method when the number of features is large.

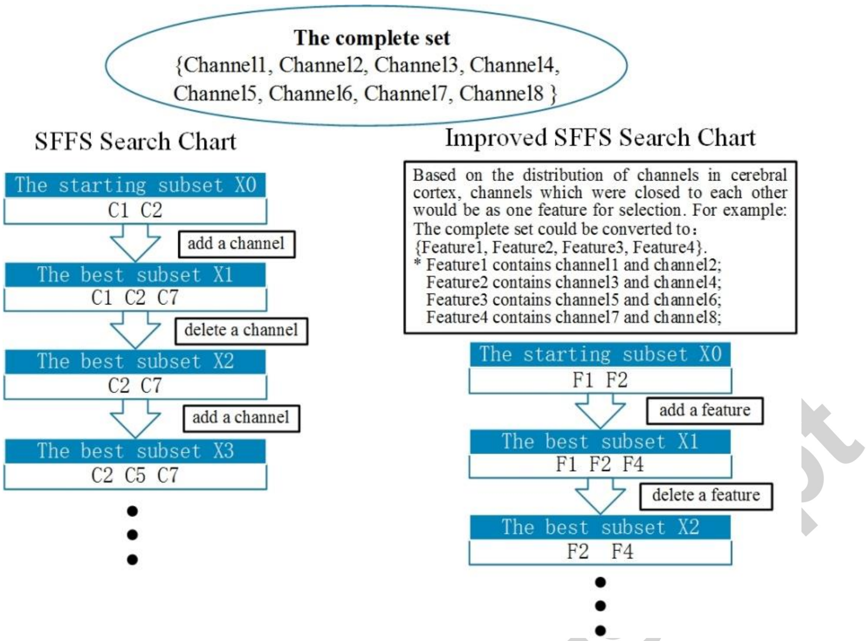


Fig. 4. The search chart of the SFFS and improved SFFS methods. The complete set of the improved SFFS contains fewer features. This improvement could save on search time. In the figure, "C" represents "channel" and "F" represents "feature".

Improvement: According to the distribution of channels in the cerebral cortex, the adjacent channels could be treated as one feature for selection. Thus, the complete set contains fewer features, and the improved SFFS could select or delete several channels each time. Fig. 4 shows the main difference between SFFS and improved SFFS methods. As shown in this figure, the complete set of improved SFFS contains fewer features. The search time could be significantly reduced.

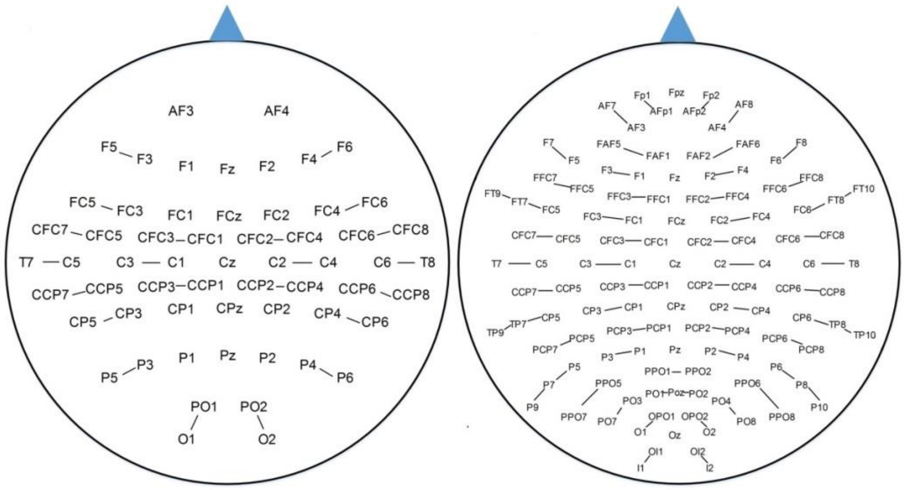


Fig. 5. The position of EEG electrodes used for data acquisition (59 channels for data 1 and 118 channels for data 2). Black lines connect the channels that are close to each other; these connected channels are treated as one feature for selection.

The channel distributions of two datasets are shown in Fig. 5. For data 1 (59 channels), the number of features in the universal set was reduced to 37. For data 2 (118 channels), the number of features in the universal set was reduced to 59.

2.5. Data Processing

The EEG data after the visual cue were used (4–7 s for data 1 and 0–3.5 s for data 2). The EEG data were band-pass filtered using a fifth-order Butterworth band-pass filter from 8 Hz to 30 Hz because this frequency band includes the range of frequencies that are mainly involved in performing MI. Then, the filtered EEG data from the training data were used to select the optimal channels. In the present study, the improved SFFS started from X_0 (channels AF3 and AF4 for data 1 and channels FP1 and AFP1 for data 2) and performed the loop of channel selection for CSP continuously. The variance of the spatially filtered signals were applied as the inputs of the SVM classifier. Tenfold cross-validation accuracy of the training data presented the performance of each channel set.

The structure of the channel selection method proposed in the present study is shown in Fig. 6. The improved SFFS performed the loop of channel selection continuously until the stop criterion was fulfilled. The CSP was used for feature extraction from the selected channels. Then, the SVM was adopted to classify the features.

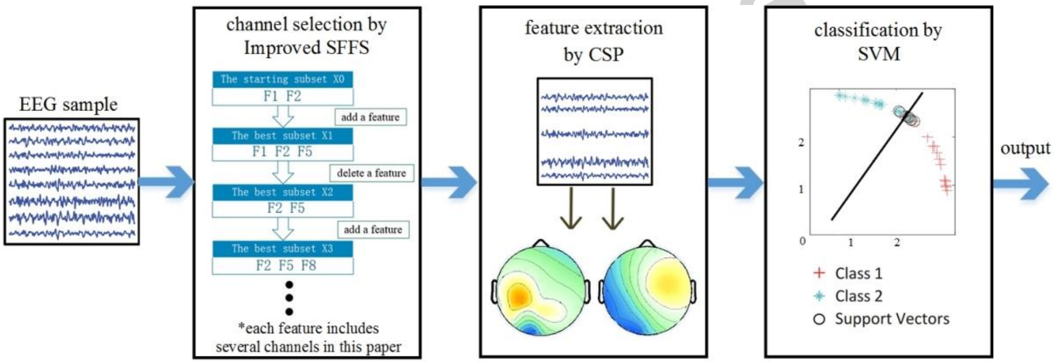


Fig. 6. The structure of the proposed method in this study.

3. Results

3.1. Channel selection

In this section, the cross-validation accuracy behavior over varying numbers of channels are presented to show how the channel number affects the performance. For each dataset, the improved SFFS performed channel selection until all the channels were used. The overall accuracy behavior averaged from all the subjects in each dataset is shown in Fig. 7. For each dataset, the accuracy initially reached a peak at a certain point and then decreased with the increase in channel number. The performances of data 1 decreased more quickly and steeply than that of data 2.

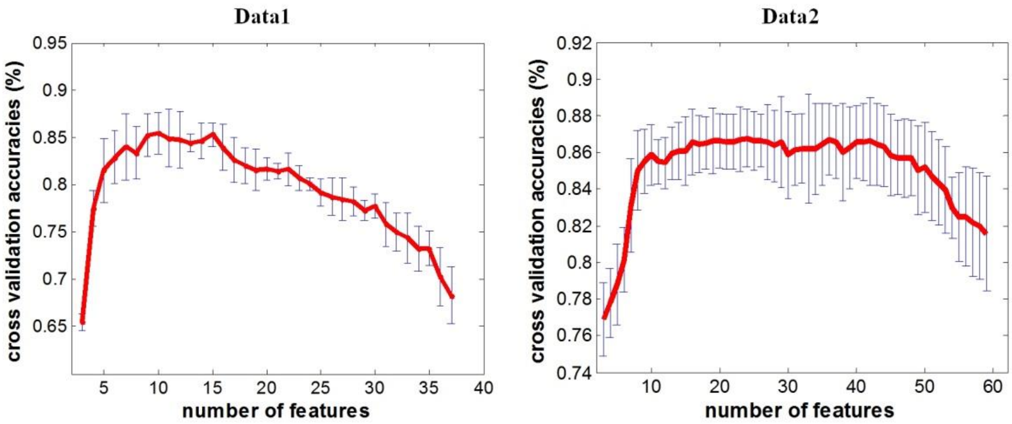


Fig. 7. Overall accuracy curves showing the accuracy behavior with the varying numbers of channels. For data 1, 59 channels were reduced to 37 features (each feature contains 1 or 2 channels). For data 2, 118 channels were reduced to 59 features (each feature contains 1 to 3 channels). The overall accuracy curves are the average over all subjects in each dataset. The red line denotes the mean validation accuracy curve, and the black vertical lines represent the envelopes of ±standard deviation.

3.2. Computation time comparison

The computation time of the two methods when cross-validation accuracy reached the optimal point are shown in Figs. 8 and 9 to compare their operating speeds.

Operating platform: 1) Hardware: processor: Intel(R) Core(TM) i5-2450 CPU @ 2.5 GHz; RAM: 8 GB; 2) Software: Windows 7 professional; Matlab R2014a.

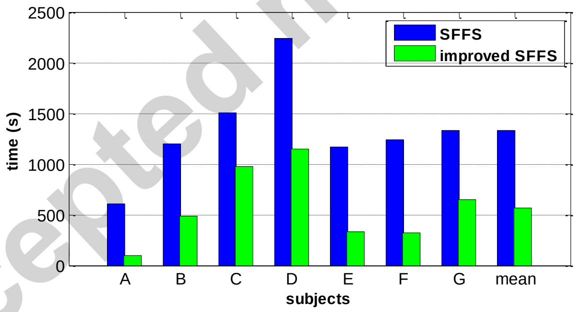


Fig. 8. Computation time comparisons between SFFS and improved SFFS for data 1.

The computation time of the improved SFFS was significantly less than that of the SFFS ($p < 0.005$) for data 1. The average computation time was reduced by 57% (760 s). Fig. 9 shows the results for data 2, which has a large number of channels. On average, the improved SFFS method reduced the computation time of the original SFFS by 65% (4,745 s). The computation time decreased significantly ($p < 0.005$) for the improved SFFS of data 2.

Comparisons between Figs. 8 and 9 reveal that the improved SFFS significantly reduced computation time when cross-validation accuracies reached the peak for both datasets. The reduction of time was more obvious for the dataset with a larger number of channels (65% reduction for data 2 and 57% reduction for data 1).

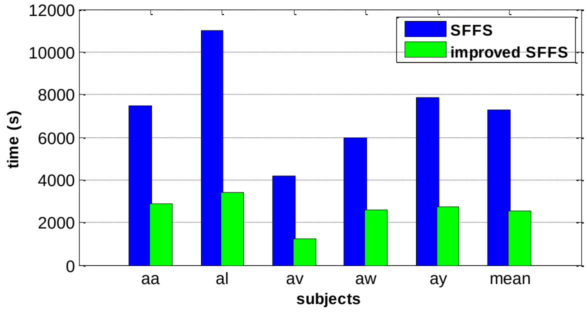


Fig. 9. Computation time comparisons between SFFS and improved SFFS for data 2.

3.3. Feature extraction

Fig. 10 presents the topographic maps from two channel configurations of subject E, aa, and ay. The ERD phenomenon mainly occurred in the left cerebral cortex area during imaging of right hand movement. The ERD phenomenon of foot MI mainly localized in the area between both hemispheres and was slightly behind the central zone.

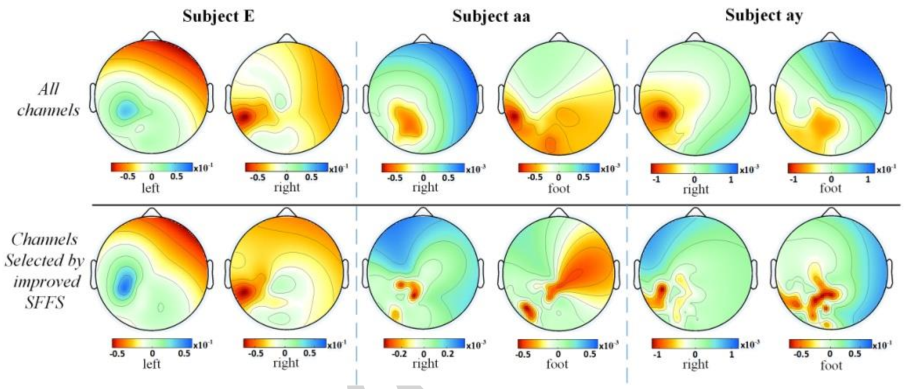


Fig. 10. Topographic maps from 2 channel configurations of subject E, aa and ay. Each topography was averaged over all trials of training data. For the no channel area of maps in the second row, the value was set to zero.

The topographic maps formed by the improved SFFS method also showed differences between two classes of MI tasks. It was almost consistent with the first row in the case of fewer channels. Selecting the appropriate channels was beneficial to feature extraction [30, 40] because it could provide effective features and remove several inconsequential channels at the same time.

3.4. Test results comparison

Tables 1 and 2 summarize the performance (classification accuracy and number of selected channels) of all subjects in the test data. Three channel selection methods were compared. The last row of the tables presents the p values obtained from the paired *t* test of the results of all channels and channels selected by other methods.

SVM-RFE is a feature selection method using weight magnitude as the ranking criterion based on SVM concept [41]. This method eliminates the feature with the smallest ranking criterion. In the present study, SVM-RFE was used to select channels and then compared with the proposed method.

For each subject, all the channels ranked by SVM-RFE and the first r channels that obtained the best cross-validation accuracy were selected.

Table 1

Performance comparison of different methods applied on data 1 with 59 channels overall.

Subject	Data 1 (BCI Competition IV datasets 1)						
	All channels	**Improved SFFS**		**SFFS**		**SVM-RFE**	
	Acc (%)	Acc (%)	Num	Acc (%)	Num	Acc (%)	Num
A	43	69	6	60	9	57	4
B	42	63	15	66	19	54	2
C	62	87	26	91	21	84	32
D	81	94	29	94	34	88	26
E	91	96	19	96	21	95	10
F	49	65	8	58	19	58	4
G	66	72	22	83	21	71	18
Mean	62	78.0	17.9	78.3	20.6	72.5	13.7
STd	18.9	14.0	8.7	16.5	7.3	16.6	11.8
p value	–	0.0025	–	0.002	–	0.0045	–

p value denotes the paired t test between classification results of all channels and other methods. Acc: classification accuracy, Num: the number of selected channels.

Table 1 shows the results for data 1. The proposed improved SFFS algorithm yielded an average classification accuracy of 78%. The improvement in the classification accuracy of the improved SFFS was substantial compared with using all the channels and shows that the improved SFFS is capable of selecting important channels. On average, the improved SFFS achieved significantly better classification accuracy than the SVM-RFE (p value = 0.012). In addition, the accuracy performances of the improved SFFS and SFFS were almost the same (p value = 0.91).

Table 2

Performance comparison of different methods applied on data 2 with 118 channels overall.

Subject	Data 2 (BCI Competition III datasets IVa)						
	All channels	**Improved SFFS**		**SFFS**		**SVM-RFE**	
	Acc (%)	Acc (%)	Num	Acc (%)	Num	Acc (%)	Num
aa	75.7	76.4	27	78.3	26	68.6	7
al	85.7	94.3	47	93.6	43	92.9	8
av	62.1	65	18	68.6	18	62.9	14
aw	87.1	89.5	27	87.9	20	80	35
ay	87.1	91.4	35	92.7	35	88.6	18
Mean	79.5	83.3	30.8	84.2	28.4	78.6	16.4
STd	10.9	12.3	10.9	10.6	10.5	12.8	11.3
p value	–	0.047	–	0.023	–	0.75	–

p value denotes the paired t test between classification results of all channels and other methods. Acc: classification accuracy, Num: the number of selected channels.

Table 2 shows the results for a large number of channels. The average classification accuracy of the improved SFFS was 83.3% and yielded an average improvement of 3.8% with the use of only 30.8 out of 118 channels. On average, the improved SFFS achieved significantly better classification accuracy than the SVM-RFE (p value = 0.045). In addition, no statistically significant differences exist between the improved SFFS and SFFS (p value = 0.39).

Tables 1 and 2 reveal that, in both datasets, the improved SFFS can significantly increase classification accuracy compared with using all the channels. Moreover, the improved SFFS achieved almost the same accuracy as the SFFS and, at the same time, significantly reduced computation time. Although the SVM-RFE was the best method to reduce the number of channels, the classification accuracy was lower than that of the improved SFFS.

4. Discussion

4.1. Differences between SFFS and SVM-RFE

The improved SFFS yielded higher overall classification accuracy than the SVM-RFE. Several differences were observed between these two methods. The improved SFFS performed channel selection from a small subset and added channels and mainly focused on the performance of the entire channel configuration rather than individual channels. The SVM-RFE method removed the channel with the lowest score at each iteration step and mainly focused on the performance of individual channels rather than the channel configuration. Thus, the SVM-RFE performed channel removal rather than channel selection [20]. The ranking procedure was suboptimal because a channel that has been eliminated was not revisited. Moreover, good individual features do not necessarily lead to good classification performance [22]. In terms of computation time, the SVM-RFE was a fast method, whereas the SFFS was a time-consuming method. These two methods were compared in the present study because of these differences.

4.2. Distributions of the selected channels

The distributions of channels selected by the improved SFFS are shown in Fig. 11. The selected channels were marked with different colors according to the number of times each channel was selected. The darker color represents a larger number.

For data 1, C3 and C4 were selected more than four times. C3 and C4 were also verified as important in right and left hand MI tasks [26, 27]. In addition, several channels besides C3 and C4 were also selected (C5, C1, C2, CCP3, and CCP4); they were all distributed in motor areas of the cerebral cortex. For data 2, five subjects all performed right hand and foot MI tasks. The distributions of generic channels were compared with topographic maps. In Fig. 10, the ERD phenomenon mainly occurred in the left cerebral cortex during right hand MI task. As shown in Fig. 11, C3 and several channels (C1, CCP3, CCP5, and CP3) around it were selected; they were all distributed in motor areas of the left cerebral cortex. The ERD phenomenon of foot imagery localized close to the primary foot area between both hemispheres [26]. In Fig 11, Cz, CCP1, and CFC2 were all distributed in this area.

Many channels distributed in the front area of the cerebral cortex were selected. These channels were unrelated to MI because the improved SFFS method used in this study selected channels according to the order from top to bottom. Thus, channels distributed in the front area of the cerebral cortex were prioritized when the cross-validation accuracies were the same. The performance of each channel could only be judged by the evaluation criterion. Thus, the selected channels were distributed in an irregular manner. However, several important channels related to MI were selected.

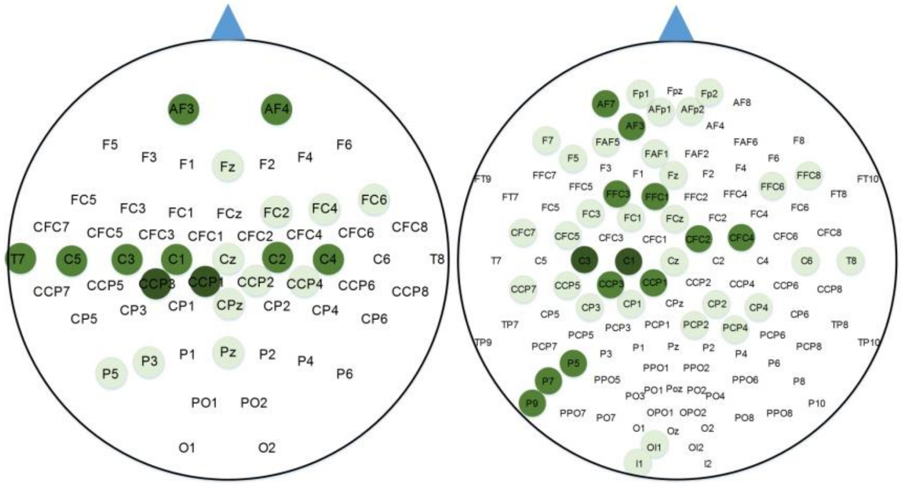


Fig. 11. The distributions of the selected channels for two datasets. For data 1, channels selected more than three times out of seven subjects were marked. CCP3 and CCP1 were selected six times. For data 2, channels selected more than one time out of five subjects were marked. C1 and C3 were selected four times.

5. Conclusion

The SFFS method has been considered one of the best feature selection methods [54]. However, the SFFS was time consuming, particularly when the number of features was large. In this study, the improved SFFS method was introduced to select channels for CSP in MI-based BCI. Based on the distribution of channels in the cerebral cortex, a strategy for simultaneous selection of multiple channels was implemented. Experimental studies on two public EEG datasets (BCI Competition IV datasets 1 and BCI Competition III datasets IVa) indicated that the improved SFFS could significantly reduce the computation time. The improved SFFS can also obtain higher classification accuracy than the SVM-RFE.

In summary, this study mainly provided a way to optimize EEG channels in terms of reducing computation time. This study presented two major contributions. First, the improvement significantly reduced the computation time of the SFFS algorithm without compromising the classification accuracy. Reducing the time of channel selection is good for the practical application of BCIs. Second, this study introduced an idea for channel selection, which combines the actual distribution of channels and the intelligent selection algorithm. This combination might lead to good results for feature selection. In future studies, the proposed method could be used in more feature selection cases and should be evaluated using more EEG data.

Acknowledgements

This work was supported in part by the Grant National Natural Science Foundation of China, under Grant Nos. 61573142, 61203127, 91420302 and 61305028. This work was also supported by the

Fundamental Research Funds for the Central Universities (WG1414005, WH1314023, and WH1516018) and Shanghai Chenguang Program under Grant 14CG31.

Reference

[1] I. Daly, S.J. Nasuto, K. Warwick, Single tap identification for fast BCI control, Cogn Neurodyn. 5(1) (2011) 21–30

[2] C.T. Lin, Y.K. Wang, S.A. Chen, An EEG-based brain–computer interface for dual task driving detection, Neurocomputing.129(4) (2014):85–93.

[3] J.R. Wolpawa, N. Birbaumerc, D.J. McFarlanda, G. Pfurtscheller, T.M. Vaughana, Brain-computer interfaces for communication and control, Clin Neurophysiol. 113(6) (2002) 767-791.

[4] J. Jin, B.Z Allison, E.W Sellers, C. Brunner, P. Horki, X.Y. Wang, C. Neuper, An adaptive P300-based control system, J Neural Eng. 8(3) (2011) 292-301.

[5] J. Jin, I. Daly, Y. Zhang, X.Y. Wang, A. Cichocki. An optimized ERP brain–computer interface based on facial expression changes, J Neural Eng. 11(3) (2014) 1082-1088.

[6] L. Chen, J. Jin, Y. Zhang, X.Y. Wang, A. Cichocki, A survey of the dummy face and human face stimuli used in BCI paradigm, J Neurosci Methods. 239 (2015) 18-27.

[7] J.H. Pan, Y.Q. Li, Z.H. Gu, Z.L. Yu, A comparison study of two P300 speller paradigms for brain–computer interface, Cogn Neurodyn. 7(6) (2013) 523-529.

[8] T. Kaufmann, S. M. Schulz, A. Köblitz, G. Renner, C. Wessig, A. Kübler, Face stimuli effectively prevent brain-computer interface inefficiency in patients with neurodegenerative disease, Clinical Neurophysiology Official Journal of the International Federation of Clinical Neurophysiology, 124(5) (2013) 893–900.

[9] J. Jin, E.W. Sellers, S.Zhou, Y. Zhang, X. Wang, A. Cichocki, A P300 brain computer interface based on a modification of the mismatch negative paradigm, International Journal of Neural Systems, 25(3) (2015) 595-599.

[10] E. Yin, T. Zeyl, R. Saab, D. Hu, Z. Zhou, T. Chau, An auditory/tactile visual saccade-independent P300 brain-computer interface, International Journal of Neural Systems, 26(1) (2015)1650001.

[11] J.N.D. Cruz, F. Wan, C. M. Wong, T. Cao, Adaptive time-window length based on online performance measurement in SSVEP-based BCIs, Neurocomputing, 149(PA) (2015) 93-99.

[12] Y.S. Zhang, D.Q. Guo, K.W. Cheng, D.Z. Yao, P. Xu, The graph theoretical analysis of the SSVEP harmonic response networks, Cogn Neurodyn. 9(3) (2015) 305-315.

[13] G. Pfurtscheller, Graphical display and statistical evaluation of event-related desynchronization (ERD), Electroencephalogr Clin Neurophysiol. 43(5) (1977) 757-760.

[14] G. Pfurtscheller, Event-related synchronization (ERS) an electrophysiological correlate of cortical areas at rest, Electroencephalogr Clin Neurophysiol. 83(1) (1992) 62-69.

[15] T. Yu, Y. Li, J. Long, Z. Gu, Surfing the internet with a BCI mouse, Journal of Neural Engineering, 9(3) (2012) 36012-36025(14).

[16] F. Velasco-Álvarez, R. Ron-Angevin, L. D. Silva-Sauer, Salvador-Ros, Audio-cued motor imagery-based brain-computer interface: Navigation through virtual and real environments, Neurocomputing, 121(18) (2013) 89-98.

[17] B. Blankertz, F. Losch, M. Krauledat, G. Dornhege, G. Curio, K.R. Muller, The Berlin brain-computer interface: Accurate performance from first-session in BCI-naive subjects, IEEE Trans. Biomed. Eng. 55(20) (2008) 2452–2462.

[18] M. Arvaneh, C.T. Guan, K.K. Ang, C. Quek, Optimizing the channel selection and classification accuracy in EEG based BCI, IEEE Trans Biomed Eng. 58(6) (2014) 1865 - 1873.

[19] F. Popescu, S. Fazli, B. Yakob, B. Blankertz, K.R. Muller, Single trial classification of motor imagination using 6 dry EEG electrodes, PLoS ONE. 2 (2007) 1–5.

[20] T.N. Lal, M. Schroder, T. Hinterberger, J. Weston, M. Bogdan, N. Birbaumer, B. Scholkopf, Support vector channel selection in BCI, IEEE Trans Biomed Eng. 51(6) (2004) 1003-1010.

[21] A. Al-Ani, A. Al-Sukker, Effect of feature and channel selection on EEG classification, Conf Proc IEEE Eng Med Biol Soc. (2006) 2171-2174.

[22] C.Y. Kee, S.G. Ponnambalam, C.K. Loo, Multi-objective genetic algorithm as channel selection method for P300 and motor imagery data set. Neurocomputing. 161 (2015) 120-131.

[23] L. He, Y. Hu, Y. Li, Channel selection by Rayleigh coefficient maximization based genetic algorithm for classifying single-trial motor imagery EEG, Neurocomputing. 121(18) (2013) 423-433.

[24] Y. J. Wang, S. Gao, X.R. Gao, Common spatial pattern method for channel selection in motor imagery based brain-computer interface, Proc 27th Annu Int Conf Eng Med Biol Soc. (2005) 5392 -5395.

[25] J. Lv, M.C. Liu, Common spatial pattern and particle swarm optimization for channel selection in BCI, ICICIC IEEE Computer Soc. (2008) 457.

[26] G. Pfurtscheller, F.H.L.D. Silva Event-related EEG/MEG synchronization and desynchronization: basic principles. Clin Neurophysiol. 110(11) (1999) 1842-1857.

[27] S. Ge, R.M. Wang, D.C. Yu, Classification of Four-Class Motor Imagery Employing Single-Channel Electroencephalography, PLoS ONE. 9(6) (2014) 1-7.

[28] G. Pfurtscheller, C. Brunner, A. Schlögl, F.H.L.D. Silva, Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks, Neuroimage. 31(1) (2006) 153-159.

[29] M. Pregenzer, G. Pfurtscheller, Frequency component selection for an EEG-based brain to computer interface, IEEE Trans Rehabil Eng. 7(4) (1999) 413-419.

[30] H. J. Yang, C.T. Guan, K.K. Ang, S. Kok, C. Phua, C.C. Wang, Selection of effective EEG channels in brain computer interfaces based on inconsistencies of classifiers, Conf Proc IEEE Eng Med Biol Soc. (2014) 672-675.

[31] I. Guyon, A. Elisseeff, An Introduction to Variable and Feature Selection, Journal of Machine Learning Research, 3 (2003)1157-1182.

[32] T. Lan, D. Erdogmus, A. Adami, M. Pavel, S. Mathan, Salient EEG channel selection in brain computer interfaces by mutual information maximization, in Proc. 27th Annu. Int. Conf. Eng. Med. Biol. Soc. (2005) 7064–7067.

[33] T. Lan, D. Erdogmus, A. Adami, S. Mathan, M. Pavel, Channel selection and feature projection for cognitive load estimation using ambulatory EEG, Comput. Intell. Neurosci. (2007) 8-20.

[34] L.J. Meng, J. Jin, X.Y. Wang, A comparison of three electrode channels selection methods applied to SSVEP BCI, BMEI 4th Int Conf on IEEE. (2011) 584-587.

[35] H. Peng, F. Long, C. Ding, Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy, IEEE Trans. Pattern Anal. Mach. Intell. 27 (8) (2005) 1226–1238.

[36] H. Ramoser, J. Muller-Gerking, G. Pfurtscheller, Optimal spatial filtering of single trial EEG during imagined hand movement, IEEE Trans Rehabil Eng. 8(4) (2000) 441-446.

[37] G. Huang, G.Q. Liu, J.J. Meng, D.G. Zhang, X.Y. Zhu, Model based generalization analysis of common spatial pattern in brain computer interfaces, Cogn Neurodyn. 4(3) (2010) 217-223.

[38] J. Meng, G. Huang, D. Zhang, X. Zhu, Optimizing spatial spectral patterns jointly with channel configuration for brain–computer interface, Neurocomputing, 104 (2013) 115-126.

[39] Y. Zhang, G. Zhou, J. Jin, X. Wang, A. Cichocki, Optimizing spatial patterns with sparse filter bands for motor-imagery based brain–computer interface, J Neurosci Methods, 255 (2015) 85-91.

[40] M. Arvaneh, C. Guan, K.K. Ang, C. Quek, Optimizing the channel selection and classification accuracy in EEG-based BCI, IEEE Trans. Bio-med. Eng. 58 (6) (2011) 1865–1873.

[41] F. Goksu, N.F. Ince, A.H. Tewfik, Greedy solutions for the construction of sparse spatial and spatio-spectral filters in brain–computer interface applications, Neurocomputing, 108 (2013) 69–78.

[42] X. Yong, R. K. Ward, G. E. Birch, Sparse spatial filter optimization for EEG channel reduction in brain-computer interface, IEEE ICASSP'08, (2008) 417-420.

[43] P. Pudil, J. Novovičová, J. Kittler, Floating search methods in feature selection, Pattern recognition letters. 15(11) (1994) 1119-1125.

[44] I. Guyon, J. Weston, S. Barnhill, V. Vapnik, Gene selection for cancer classification using support vector machines, Mach Learn. 46(1-3) (2002) 389-422.

[45] L. F. Nicolas-Alonso, R. Corralejo, J. Gomez-Pilar, D. Álvarez, R. Hornero, Adaptive semi-supervised classification to reduce intersession non-stationarity in multiclass motor imagery-based brain computer interfaces. Neurocomputing, 159(1) (2015) 186-196.

[46] L. He, Z.H. Gu, Y.Q. Li, Z.L. Yu, Classifying Motor Imagery EEG Signals by Iterative Channel Elimination according to Compound Weight, Artificial Intelligence and Computational Intelligence Springer Berlin Heidelberg. (2010) 71-78.

[47] B. Nasihatkon, R. Boostani, M.Z. Jahromi, An efficient hybrid linear and kernel CSP approach for EEG feature extraction, Neurocomputing. 73(1-3) (2009) 432-437.

[48] K.K. Ang, Z.Y. Chin, H.H. Zhang, C.T. Guan, Robust filter bank common spatial pattern (RFBCSP) in motor-imagery-based brain-computer interface, Conf Proc IEEE Eng Med Biol Soc. (2009) 578-581.

[49] C. Liu, H.B. Zhao, C.S. Li, H. Wang, Classification of ECoG motor imagery tasks based on CSP and SVM, BMEI 3rd Int Conf on IEEE. (2010) 804-807.

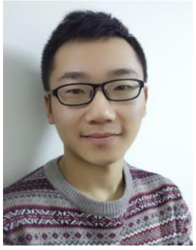
[50] A. M. Alvarez-Meza, L. F. Velasquez-Martinez, G. Castellanos-Dominguez, Time-series discrimination using feature relevance analysis in motor imagery classification, Neurocomputing, 151(P1) (2015) 122-129.

[51] V.N. Vapnik, V. Vapnik, Statistical learning theory, New York Wiley (1998).

[52] J. Nocedal, S.J. Wright, Numerical Optimization, Springer New York (2006).

[53] Y. Yang, J. Wang, Y.Y. Yang, Improving SVM classifier with prior knowledge in micro calcification detection, Conf Proc IEEE Ima Pro. (2012) 2837-2840.

[54] Y. Peng, Z. Wu, J. Jiang, A novel feature selection approach for biomedical data classification, Journal of Biomedical Informatics, 43(43) (2010) 15-23.



Zhaoyang Qiu received the bachelor of engineering in control science and engineering from the East China University of Science and Technology, Shanghai, China, in 2010. He is currently pursuing his Ph.D. degree in the same university. His research interest includes brain-computer interface and signal processing.

Jing Jin received the Ph.D. degree in control theory and control engineering from the East China University of Science and Technology, Shanghai, China, in 2010. His Ph.D. advisors were Prof. Gert Pfurtscheller at Graz University of Technology from 2008 to 2010 and Prof. Xingyu Wang at East China University of Science and Technology from 2006 to 2008. He is currently a Professor at East China University of Science and Technology. His research interests include brain-computer interface, signal processing and pattern recognition.

Hak-Keung Lam received the B.Eng. with honors and Ph.D. degrees from the Department of Electronic and Information Engineering, Hong Kong Polytechnic University, Hung Hom, Hong Kong, in 1995 and 2000, respectively. During the period from 2000 to 2005, he worked with the Department of Electronic and Information Engineering, Hong Kong Polytechnic University, Hung Hom, Hong Kong, as a Postdoctoral and Research Fellow. In 2005, he joined King's College London, London, UK, as a Lecturer and is currently a Senior. He is an IEEE Senior Member. His current research interests include intelligent control and computational intelligence. In the area of intelligent control systems, he has researched fuzzy control systems, especially stability analysis, control synthesis, and performance design. In the area of computational intelligence, he works on system analysis and topology design of neural networks, fuzzy logic systems and support vector machines, random-based searching algorithms, and their applications dealing with forecasting, classification, and recognition problems. Also, he has devoted effort to the analysis and development of random-based searching algorithms such as genetic algorithm and particle swarm optimization.

Yu Zhang received the Ph.D. degree in control science and engineering from the School of Information Science and Engineering, East China University of Science and Technology, Shanghai, China, in 2013. He worked as an International Program Associate (IPA) for two years (from Dec. 2010 to Nov. 2012) in the Laboratory for Advanced Brain Signal Processing (LABSP) at RIKEN Brain Science Institute, Wako-shi, Japan. He is currently working as an Assistant Professor at the School of Information Science and Engineering, East China University of Science and Technology, Shanghai, China. His research interests include brain-computer interface, signal processing, tensor analysis, machine learning and pattern recognition.

Xingyu Wang was born in Sichuan, China, in 1944. He received the B.S. degree in mathematics from Fudan University, Shanghai, China, in 1967, and the M.S. in control theory from East China Normal University, Shanghai, China, in 1982, and Ph.D. degrees in industrial automation from East China University of Science and Technology, Shanghai, China, in 1984. He is currently a Professor at the School of Information Science and Engineering, East China University of Science and Technology, Shanghai, China. His research interests include control theory, control techniques, the application to biomedical system, and brain control.

Andrzej Cichocki received the M.Sc. (with honors), Ph.D. and Dr.Sc. (Habilitation) degrees, all in electrical engineering from Warsaw University of Technology (Poland). He spent several years at University Erlangen-Nuerenberg (Germany) as an Alexandervon-Humboldt Research Fellow and Guest Professor. In 1995-1997 he was a team leader of the Laboratory for Artificial Brain Systems, at Frontier Research Program RIKEN (Japan). He is currently a Senior Team Leader and Head of the laboratory for Advanced Brain Signal Processing, at RIKEN Brain Science Institute (Japan). He has given keynote and tutorial talks at international conferences in Computational Intelligence and Signal Processing and served as member of program and technical committees (EUSIPCO, IJCNN, ICA, ISNN, ICONIP, ICAISC), He is author of more than 300 technical journal papers and 4 monographs in English (two of them translated to Chinese. He has served as an Associated Editor of IEEE Transactions on Neural Networks, IEEE Transactions on Signals Processing, IEEE Trans of Cybernetics, Journal of Neuroscience Methods and as founding Editor in Chief for Journal Computational Intelligence and Neuroscience. Currently, his research focus on tensor decompositions, multiway blind sources separation, brain machine interface, EEG hyper-scanning. His publications currently report over 25,000 citations according to Google Scholar.